

EXAMPLE Data Management and Sharing Plan for Human Subjects Research

Acknowledgement: This document was created by the Research Data Service at the University of Illinois at Urbana-Champaign (UIUC) to provide an example Data Management and Sharing Plan (“Plan”) for an NIH proposal. We based this Plan off an award-winning DMP for [NSF- SBE “Ethnic Diversity and Public Goods Provision Across Latin America”](#) by Christopher Hale at the University of Alabama, who kindly released their DMP for reuse. We have modified this plan to reflect the requirements of NIH and resources at UIUC. **This example is provided for illustrative purposes only since all Plans submitted to NIH must reflect the specific research project proposed.** See more information here: <https://go.illinois.edu/newnih>

Project Title: Ethnic Diversity and Public Goods Provision Across Latin America

1. Data Type

Collected qualitative data will include semi-structured interviews from an estimated 50 to 70 local and regional government officials, high-ranking Catholic officials, Catholic laity, social movement organization leaders, and so on. Each interview is expected to last approximately forty minutes in length, on average. The themes of the interviews are not expected to put subjects at risk. However, the interviews will cover political activism. Such activism has made individuals potential targets for violent retribution in the past. Accordingly, all interview data will go through a de-identification process as detailed below. The qualitative interviews will be collected as digital audio files via digital voice recorder and subsequently transcribed into Microsoft Word documents.

Quantitative data will consist of two originally constructed datasets. One consists of 425 country-level time-series observations of 17 Latin American countries with data collected from the World Bank’s World Development Indicators, among other sources. The other consists of more than 130,000 individual-level survey responses collected from the Latin American Public Opinion Project and other sources. Quantitative data will be in tabular format.

Metadata to be made available includes interview questionnaires, instructions to respondents, informed consent forms, and de-identifying procedures, all available in Microsoft Word format and exportable as rich-text (TXT) format. It also includes the Stata code used to run the statistical analyses and the Stata code used to construct the datasets, for replication purposes. These files can also be converted into rich-text format.

2. Related Tools, Software and/or Code

All statistical data will be stored in Stata format (.dta), which can be easily exported into a variety of other highly accessible data formats (e.g., CSV).

2. Standards

A single folder will be created for all data collected in connection with this project. Subfolders will then be created for the quantitative datasets, metadata containing qualitative interview forms, IRB protocols, and so on, and finally an encrypted file containing two subfolders: one subfolder will contain the audio files themselves, while another subfolder will contain the de-identified interview transcriptions.

Informed consent is of vital importance in conducting social research, as is the protection of the identity of respondents. This project will soon be under review with the University’s IRB. All personnel involved in the project will have been trained to be sensitive to the ethics of research with

human subjects prior to IRB approval. Prior to interviews, a written statement will be read to respondents to inform them of the nature of the research. They may then make an informed choice about whether they are willing to participate. Respondents will also be advised that, while data (including de-identified transcripts) will be made publicly available, personal and identifying information will remain confidential. A waiver of signed consent is being requested from the IRB in favor of verbal consent to protect participant de-identification.

3. Data Preservation, Access, and Associated Timeline

De-identified transcripts as well as survey data and all metadata in connection with this project will be hosted, curated, and made freely available for research and teaching via the institutional data repository Illinois Data Bank, which is operated and managed by the Research Data Service at the University Library, University of Illinois at Urbana-Champaign. The Illinois Data Bank provides metadata, persistent identifiers (i.e., DOIs), professional curation and long-term access. Additionally, any reports, presentations, manuscripts, and other documents that record research outputs generated under this project can be deposited in IDEALS, the Illinois Digital Environment for Access to Learning and Scholarship. Both repositories are optimized for their respective content types and support robust indexing and stable access.

As outlined in NIH's Data Management and Sharing Policy, data will be shared on publication of results in a journal article or after project completion, whichever comes earlier.

4. Access, Distribution, or Reuse Considerations

A de-identification process will be developed prior to field work so that data can be de-identified as it is collected. Names will be replaced with demographic information and dates referring to specific potentially identifying events will be replaced by general markers. Prior to public dissemination, all interview data files will be solely kept in encrypted, password-protected folders on a laptop and U of I Box, a secure cloud-based storage that licensed by the University of Illinois. As soon as data is collected in each region, any lists, names, or identifying information will be permanently deleted or destroyed by shredding (if physical copies exist). The digital audio files of the interviews will be permanently deleted after five years.

5. Oversight of Data Management and Sharing

The Project PI will control all matters in connection with this project related to data collection and publication. Curation and long-term storage and preservation of the data will be handled by the Illinois Data Bank, as detailed above.